A SYSTEMATIC LITERATURE REVIEW ON AUTOIMMUNE THYROID DISEASE SUSCEPTIBILITY AND PROGNOSTIC

Nagavali Saka[1] and S. Murali Krishna[2]

[1] Research Scholar, JNTUA, Anantapur, Andhra Pradesh, India.
Email: vali214@gmail.com

[2] Professor, Department of Computer Science and Engineering,
Sri Venkateswara College of Engineering(SVCE), Tirupati, Andhra Pradesh, India.
Email: muralikrishna.s@svcolleges.edu.in

Abstract

Recently, in the human community, there has been a significant need to get awareness about the immune system in order to get away from all types of diseases. Because the Immune system is playing a major role in provides the human body with a high level of protection from invading pathogens, in a robust, self-organised and distributed manner. Auto Immune Thyroid Disease (AITD) is a severe disease in the medical domain that has to be identified as quickly and continuous treatment is required in order to reduce the side effects and consequence of this disease. As AITD is mainly affecting women, so in the perception of women's health, now has turned into a key concern among the emergent countries because of the deteriorating quality of life due to autoimmune diseases. Diagnosis of AITD is important to avoid the severity level of disease, so the symptoms, causes, susceptibility (influenced) and prognostic of the disease and recommended treatments is discussed in this work. Ultimately the objective of this work is to present the comprehensive literature review on Autoimmune Thyroid disease susceptibility (influenced) and prognostic (prediction) proposed by different researchers using data mining classification schemes, employing ensemble machine learning techniques to extend the human's life.

Key words: *Health, Death rate, Immunity, Immune system, Autoimmune diseases, Susceptibility, Prognostic, Data mining and Machine learning.*

1. Introduction

Autoimmune diseases are a heterogeneous group of chronic diseases which occur secondary to loss of self-antigen tolerance. The etiopathogenesis of autoimmune disease is uncertain. Genetic factors as well as environmental factors appear to interplay, leading to a cascade of events resulting in disease onset [17]. Autoimmune diseases are chronic, multifactorial conditions. Through machine learning (ML), a branch of the wider field of artificial intelligence, it is possible to extract patterns within patient data, and exploit these patterns to predict patient outcomes for improved clinical

management. Here, we surveyed the use of ML methods to address clinical problems in autoimmune disease [1]. Data Mining (DM) and Machine Learning (ML) play a vital role in enhancing the performance of tasks such as disease risk prediction in healthcare communities, resulting in better serving of the societies [4]. Autoimmune disease Three elements contribute to autoimmune disease development: genetic predisposition, environmental factors and immune system dysregulation (Fig. 1). Due to the heterogeneity of onset and progression, diagnosis and prognosis for autoimmune disease is unpredictable [1].

According to a survey, 20 million people are suffering from some form of thyroid disease and up to 60% of them don't know that they have thyroid disease. The number of women having thyroid disease is eight times more than men [27]. The Thyroid is a little butterfly-formed organ that is situated at the base of the neck beneath the voice box is a piece of the endocrine framework.Thyroid disease is one of the deadly disease which affect the human life worldwide. Thyroid disease is a medical condition that affects the function of the thyroid gland, which is located in the front neck of the human body[3]. Thyroid disease has become the second largest disease in the endocrine field only to diabetes, which has attracted wide attention over the world. The correct classification of the disease is an important part of clinical diagnosis [30].

2. Literature Survey on Thyroid Susceptibility

The functional thyroid disease is mainly divided into hypothyroidism (underactivethyroid) and hyperthyroidism (overactivethyroid) that is further subdivided into overt and subclinical disease [10]. Autoimmune thyroid disease is the most common form of thyroid dysfunction causing several forms of thyroiditis ranging from hypothyroidism (Hashimoto's thyroiditis) and hyperthyroidism (Graves's Disease). Thyroid autoimmunity is characterized by thyroid autoantibodies, especially anti-TPO and anti-Tg[10].Thyroid segmentation and volume reconstruction are hence essential to diagnose thyroid related diseases as most of these diseases involve a change in the shape and size of the thyroid over time [13].The improper secretion of thyroid may lead to obesity, fertility related problems, feeling depressed, etc., Most of the thyroid problems can be managed if it has been properly treated [6]. The relationship between stress and autoimmune diseases appears to be complex and intricate. A formidable amount of evidence supports the association and the role of stress in influencing various aspects of autoimmune diseases, including for instance disease onset and exacerbations[17]. Age and Sex have been considered as important features because Thyroid disorders are said to occur during a particular age range (17-54 years) and is more prevalent in females [5]. Undiagnosed or inadequately treated hypothyroidism may cause miscarriage, preterm delivery and severe developmental problems in children in pregnant women [27].

3. Literature Survey of Prognostic Methods

Disease diagnosis plays a major role and it is indispensable for any busy clinician. Thyroid disease is one such disease and prediction of which is a difficult aspect without a computer technology [28]. Prediction and diagnosis of disease play a critical role and it is indispensable at clinical level. Diagnosis of thyroid deceases is very challenging and critical issues in medical science. Data mining based classification techniques play a very important role in diagnosis of thyroid diseases [33]. Two common diseases of the thyroid gland, which releases thyroid hormones for regulating the rate of the body's metabolism, are hyperthyroidism and hypothyroidism. Classification of these thyroid diseases is a considerable task [5].Accurate and early detection of a disease is more important and compulsory in healthcare domain to facilitate correct and prompt diagnosis and timely treatment [11]. Machine learning generally involves utilization of data mining techniques and some associated learning algorithm to construct models of what is going into the reservoir of data. It provides methodology and technology to transform these heaps of data into useful information which can be used for decision making or predicting future outcomes. This information is extracted through various data mining techniques and algorithms such as association, classification, clustering, and pattern recognition and is of great use to the medical experts[12].Lots have been done at the clinic level for the effective diagnosis of thyroid, however, use of machine learning framework for the detection and diagnosis of thyroid can be achieved great results[12]. Healthcare using classification techniques brought new ways to advance the technology with automatic predictions and quick treatment of patients and it is very clear that Healthcare organization can adopt predicting technologies using classing algorithms for healthy living [15]. The complexity of multi-relational data mining put forward to higher requirements on the classification precision and the algorithm stability of medical diagnosis, so the single classifier cannot meet the requirement any more. Therefore, the ensemble methods were used to analyzethe data of thyroid disease [30]. The selection of data mining tool and the optimization algorithm will show an impact on the speed and accuracy[18].

4. Embrical Survey on Classification Schemes and Machine Learning techniques

[5]discussed three classification techniques have been used namely Naive Bayes, Support vector machines and Random Forest for the prediction of thyroid disease. They found that the Support Vector Machines are the most accurate technique with the accuracy 92.92%. They also suggested SVM classifier to separate the symptoms of thyroid diseases into 4 classes, namely Hypothyroid, Hyperthyroid, Sick Euthyroid and Euthyroid (negative).

[6] applied and found that the classifier's accuracy is increased after selecting the important features with feature identification in multiclass classification of thyroid data and conclude that prediction of any disease can be done accurately by applying classification models with feature selection.

[7] has been done on the thyroid dataset by different machine learning classifiers such as decision tree, random forest tree, extra tree, and bagging ensemble model. The seed value 35 and num-fold value 10 have found the highest accuracy using bagging ensemble techniques and concluded that the bagging ensemble technique is the best compared with the other three classifier algorithms.

[8] Neethi Priya,proposed a method to predict the thyroid disorder at earlier stage using data mining techniques helps to minimize the noise data of a patient and discussed three classification techniques have been used namely KNN, Naïve bayes, Support vector machine. The results of these classification methods are based on accuracy and performance of the model and found that the accuracy using SVM is 0.82, Naïve Bayes is 0.83 and KNN is 0.85.

[10] suggested that it may be beneficial to consider testing for anti-TPO in conjunction with the primary thyroid markers,TSH and FT4,toprevent long-term morbidity[10].

[11]has been analyzed and done the comparative study of hypothyroid disorder with four different classifiers; KNN, SVM, LR, NN. By investigating the results, it is inferred that SVM classifier is better for prediction of the level of hypothyroid disorder.They found that the hypothyroid disorder is a vital classification issue from the point of view of women's wellness. It has to be addressed and diagnosed at its early states because due to hormonal changes it varies time to time.

[12] suggests that the accuracy of prediction can be enhanced by implementing ensemble classifier rather than a straightforward classification algorithm.

[14] discussed how to predict the thyroid disorder at earlier stage using data mining techniques such as KNN, Naïve Bayes, Support vector machine, ID3 are based on speed, accuracy and performance of the model and cost for the treatment.

[15] implementedfive classification techniques for the prediction of diseases. The classification model is generated based on the training data, and the data is tested though predictions in the prediction stage. Three datasets namely Diabetics, Thyroid, Breast cancer were tested using a classification algorithm using Python environment. Classification and regression problems were solved mostly by these supervised classification algorithms. It is clear that Support Vector Machine Classification Algorithm has given the best accuracy results when compared to other techniques.

[16] suggested that data support link between maternal and child immune conditions and autism spectrum disorder (ASD), and further suggest that associations may be influenced by disease severity in the mother and ASD phenotype of the child.

[21]discussed that more attributes mean a patient has to undergo a greater number of clinical tests which is both cost effective as well time consuming andsuggested that, there is a need to develop an algorithms and thyroid disease predictive models which require minimum number of parameters of a person to diagnose thyroid disease and saves both money and time of the patient.

[23] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez , compare the different machine learning techniques on a small dataset, and tried to improve the accuracy of the aforementioned techniques in order to achieve a better comparison.

[24] Mr. Sudhir M. Gorade, Prof. Ankit Deo, Prof.Pritesh Purohit, discussed Decision tree classifiers, Bayesian classifiers, classification by back propagation, support vector machines, these techniques are eager learners they use training tuples to construct a generalized model [24].

[25] discussed a comparative analysis on diagnosis of thyroid disease by using various classification algorithms, i.e. SVM, FKNN, and decision tree. After comparing the results, found that the SVM classifier technique provides better accuracies as compared to last works. It can be inferred that SVM could be successfully used for the diagnosis of thyroid disease.

[27] proposed a hybrid system,it is comprised of feature selection process using information gain method which decreases computation time and increases the accuracy of the resulting model, k-NN Imputation for missing data values and ANFIS system, which maximize the generalization capability of our thyroid diagnosis system and results proved that proposed diagnosis system has better performance than non-hybrid schemes.

[30] proposed a new classification algorithm for thyroid diseases based on random forest is proposed. It consists in splitting the feature set into K subsets, running PCA separately on each subset and constructing a rotation matrix to generate new feature space, which improved the accuracy of the ensemble classifier, and solved the accuracy-diversity dilemma. Proved that the proposed method can achieve higher classification accuracy on the UCI standard dataset and the real clinical medical datasets with lower feature dimension and less processing time.

[31]proposed an Evolutionary Multivariate Bayisean Prediction classifier model achieves remarkable dimensionality reduction from among the 7200 medical datasets obtained from the UCI repository with 21 attributes (Continuous -15; Discrete - 6). 21 epohs (runs) are carried out for the

data and after stabilization, the data are classified as Hyper, Hypo and Normal classes. The results are evaluated based on ten evaluation metrics and the accuracy of classification is 97.97%.

[32] applied LDA data mining classification techniques is used to classify the hypothyroid disease. K-fold cross validation is also performed. The LDA Algorithm gives 99.62% accuracy with k=6 folds cross validation [32].

[33] discussed an ensemble of C4.5 and Random forest model give 99.47% of accuracy in case of information gain feature selection technique for classification of thyroid deceases [33].

[35] has been done the classification data mining techniques K nearest neighbor, Support vector machine, Decision tree and Naive Bayes for the diagnosis of thyroid disease and conclude that the Decision Tree classifier outperformed over other classifiers. They suggest that, merge it with any other classification technique such as neural network, then the result might be even better as compared to what we got with the current study.

5. Empirical Survey on Methodologies in Classification Concept

The following table summarises the various classification Methods used by previous studies are captured with our proposed study as shown in Table 1. We have selected some old research work in the overall study during this research survey. All the research work is related to thyroid and other medical data use in machine learning classifiers for prediction. Most of the members used the KNN classifier and conclude it is the best classification method.

Table 1 Summary of various Classification Methods used by the previous research

Author	Proposed Year	Reference	Method used	Accuracy
I. S. Stafford,M. Kellermann, E. Mossotto, R. M. Beattie, B. D. MacArthur and S. Ennis	2020	5	• Naive Bayes • Support Vector Machines • Random Forest	SVM(92.92 %)
S. Nandhinidevi, S. Poorani, P. Gokila Brindha.	(2020)	6	• Random Forest • KNN	KNN(96%)
Dr. Dayanand Jamkhandikar , Neethi Priya.	2020	8	• Naïve Bayes • KNN • Support Vector Machine	KNN(85%)
Gyanendra Chaubey , Dhananjay Bisen ,Siddharth Arjaria, Vibhash Yadav.	2020	9	• Logistic regression • Decision tree • KNN	KNN(87.23%)

Vaishali S. Vairale, Dr. Samiksha Shukla.	2019	11	• KNN • Support Vector Machines • Logistic regression • NN (Artificial Neural Network)	KNN(87.23%)
Umar Sidiq, Dr. Syed Mutahar Aaqib, Dr. Rafi Ahmad Khan	2019	35	• KNN • Support Vector Machines • Decision Tree • Naive Bayes	KNN(86.23%)
M. Deepika, Dr. K. Kalaiselvi.	2018	18	• SVM • Decision Tree • Artificial Neural Networks.	SVM(97.23%)
Ankita Tyagi, Retika Mihra, Aditya Saxena.	2018	21	• SVM • KNN • DT(Decision Tree)	KNN(88.23%)
M.P.Gopinath.	2017	25	• SVM • KNN • DT	SVM(96.30%) 215 -samples
G. Rasitha Banu.	2016	32	• LDA	97.62%
Suman Pandey, Anshu Tiwari , Akhilesh Kumar Shrivas ,Vivek Sharma	2015	33	• C4.5 • Random Forest	RF(96.7%)

6. Conclusion

Disease diagnosis plays a major role and it is indispensable for any busy clinician. Thyroid disease is one such disease and prediction of which is a difficult aspect without a computer technology. In this survey paper, the authors have given an elaborate work that has been done earlier using researchers using data mining classification schemes, employing ensemble machine learning techniques. Based on the usage of these techniques, the authors have tried to show the path for the feature research, some intelligent techniques such as fuzzy logic, artificial neural networks or their hybrid techniques as adaptive Neuro-fuzzy inference system can be used for classification of the thyroid disease.

7. References

[1] I. S. Stafford,M. Kellermann, E. Mossotto, R. M. Beattie, B. D. MacArthur and S. Ennis (2020).A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases. Natural Partner Journals(NPJ), npj Digital Medicine 30.

[2] Yongfeng Wang1, Wenwen Yue2, Xiaolong Li2, Shuyu Liu3, Lehang Guo2, Huixiong Xu2, Heye Zhang 4, (Member, IEEE), And Guang Yang 5, (Member, IEEE) (2020).Comparison Study of Radiomics and Deep Learning-Based Methods for Thyroid Nodules Classification Using Ultrasound Images. IEEE Access, 52010-52016.

Digital Object Identifier 10.1109/ACCESS.2020.2980290.

[3] Mrs.K.Sindhya (2020).Effective Prediction Of Hypothyroid Using Various Data Mining Techniques.EPRA International Journal of Research and Development (IJRD),311-317.

SJIF Impact Factor: 6.260| ISI I.F.Value:1.241| Journal DOI: 10.36713/epra2016ISSN: 2455-7838(Online)

[4] Syed Javeed Pasha, E. Syed Mohamed (2020), Ensemble Gain Ratio Feature Selection (EGFS) Model with Machine Learning and Data Mining Algorithms for Disease Risk Prediction. Proceedings of the Fifth International Conference on Inventive Computation Technologies (ICICT), 590-596.

[5] I. S. Stafford,M. Kellermann, E. Mossotto, R. M. Beattie, B. D. MacArthur and S. Ennis (2020).A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases. Natural Partner Journals(NPJ), npj Digital Medicine 30.

[6] S. Nandhinidevi, S. Poorani, P. Gokila Brindha (2020), Machine Learning Models for Relevant Feature Identification and Classification of Thyroid Data.International Journal of Innovative Technology and Exploring Engineering (IJITEE),1961-1963.

[7] Dhyan Chandra Yadav & Saurabh Pal(2020),Prediction of thyroid disease using decision tree ensemble method,Human-Intelligent Systems Integration https://doi.org/10.1007/s42454-020-00006-y.

[8] Dr. Dayanand Jamkhandikar , Neethi Priya (2020),Thyroid Disease Prediction Using Feature Selection And Machine Learning Classifiers,The International journal of analytical and experimental modal analysis,175-180.

[9] Gyanendra Chaubey, Dhananjay Bisen,Siddharth Arjaria, Vibhash Yadav **(2020),**Thyroid Disease Prediction Using Machine Learning Approaches,Natl. Acad. Sci. Lett. https://doi.org/10.1007/s40009-020-00979-z.

[10] ThushaniSiriwardhane, KarthikKrishn, VinodhRanganathan,VasanthJayaraman Tianhao Wang,KangBe,SarahAshman,KarenahRAjasekaran,JohnJ.Rajasekaran, HariKrishnamurthy(2019),Significance of Anti-TPO as an Early Predictive Marker in Thyroid Disease, Hindawi Autoimmune Diseases, https://doi.org/10.11,55/2019/1684074, 1-6.

[11] Vaishali S. Vairale, Dr. Samiksha Shukla (2019), Classification of Hypothyroid Disorder using Optimized SVM Method, Second International Conference on Smart Systems and Inventive Technology (ICSSIT 2019) IEEE Xplore, 258-263.

[12] Mir Saleem, S Jahangeer Sidiq,Akhtar Rasool Malik (2019), Diagnosis and Classification of Thyroid Disorder using Machine Learning - A Systematic Review, Journal of Emerging Technologies and Innovative Research (JETIR), 1040-1048.

[13] Prabal Poudel , (Member, Ieee), Alfredo Illanes , Elmer J. G. Ataide, (Member, Ieee), Nazila Esmaeili, (Member, Ieee), Sathish Balakrishnan, And Michael Friebe , (Senior Member, IEEE) (2019), Thyroid Ultrasound Texture Classification Using Autoregressive Features in Conjunction With Machine Learning Approaches. IEEE Access, 79354-79365

Digital Object Identifier 10.1109/ACCESS.2019.2923547

[14] Bibi Amina Begum,Dr.Parkavi A (2019), Prediction of thyroid Disease Using Data Mining Techniques,2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 342-345.

[15] E. Laxmi Lydia, N. Sharmil , K. Shankar and Andino Maseleno **(2019),**Analysing the Performance of Classification Algorithms on Diseases Datasets, International Journal on Emerging Technologies ,10(3)-224-230.

[16] Lisa A. Croen , Yinge Qian, Paul Ashwood, Julie L Daniels, Daniele Fallin, Diana Schendel, Laura A. Schieve , Alison B. Singer, and Ousseny Zerbo(2018),Family History of Immune Conditions and Autism Spectrum and Developmental Disorders: Findings from the Study to Explore Early Development,International Society for Autism Research(INSAR), 1-13.

[17] Kassem Sharif, Abdulla Watad, Louis Coplan, Benjamin Lichtbroun, Alec Krosser, Michael Lichtbroun, Nicola Luigi Bragazzi, Howard Amital, Arnon Afek, Yehuda Shoenfeld(2018), The role of stress in the mosaic of autoimmunity. An overlooked association, AUTREV , 1-16,doi:10.1016/j.autrev.2018.04

[18] M. Deepika,Dr. K. Kalaiselvi(2018), A Empirical study on Disease Diagnosis using Data Mining Techniques(2018),International Conference on Inventive Communication and Computational Technologies (ICICCT), IEEE Xplore, 615-620.

[19] Abeer A. Raweh, Mohammed Nassef, and Amr Badr (2018),A Hybridized Feature Selection and Extraction Approach for Enhancing Cancer Prediction based on DNA Methylation,IEEE Access, 1-13.Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

[20] Turki Turki(2018),An Empirical Study of Machine Learning Algorithms for Cancer Identification,International Conference on Networking, Sensing and Control (ICNSC),

IEEE Xplore, 978-982.

[21] Ankita Tyagi, Retika Mihra, Aditya Saxena(2018), Interactive Thyroid Disease Prediction System Using Machine Learning Technique, IEEE International Conference on Parallel, Distributed and Grid Computing(PDGC), 689-693.

[22] S.Priya, Dr.R. Manavalan(2018), Tuning the Parameters of Weighted ELM using IWO and BAT Algorithm to Improve the Classification Performance ,IEEE Xplore,547-552.

[23] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez(2017),A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease, IEEE Symposium on Computers and Communication (ISCC), 1-4.

[24] Mr. Sudhir M. Gorade, Prof. Ankit Deo , Prof.Pritesh Purohit(2017),A Study of Some Data Mining Classification Techniques,International Research Journal of Engineering and Technology (IRJET),3112-3115.

[25] M.P.Gopinath(2017), Comparative Study on Classification Algorithm for Thyroid Data Set, International Journal of Pure and Applied Mathematics,53-62.

[26] Naoki Oishi, Tetsuo Kondo Aya Ebina, Yukiko Sato, Junko Akaishi, Rumi Hino, Noriko Yamamoto, Kunio Mochizuki, Tadao Nakazawa, Hiroshi Yokomichi, Koichi Ito, Yuichi Ishikawa and Ryohei Katoh(2017), Molecular alterations of coexisting thyroid papillary carcinoma and anaplastic carcinoma: identification of TERT mutation as an independent risk factor for transformation,Modern Pathology advance online publication,1-11.

[27] Waheed Ahmad, Lican Huang, Ayaz Ahmad, Farooq Shah, Amjad Iqbal, Asma saeed(2017),Thyroid Diseases Forecasting Using a Hybrid Decision Support System Based on ANFIS, k-NN and Information Gain Method,Journal of Applied Environmental and Biological Sciences,78-85.

[28] Shaik Razia, M. R. Narasinga Rao(2016),Machine Learning Techniques for Thyroid Disease Diagnosis - A Review,Indian Journal of Science and Technology,1-9.

[29] Ebru Turanoglu-Bekar, Gozde Ulutagay, Suzan Kantarcı-Savas(2016),Classification of Thyroid Disease by Using Data Mining Models: A Comparison of Decision Tree Algorithms,Oxford Journal of Intelligent Decision and Data Science,13-28.

[30] Qiao Pan,Yuanyuan Zhang, Min Zuo, Lan Xiang, Dehua Chen(2016),Improved Ensemble Classification Method of Thyroid Disease Based on Random Forest,International Conference on Information Technology in Medicine and Education,567-571.

[31] K.Geetha & Capt. S. Santhosh Baboo(2016),An Empirical Model for Thyroid Disease Classification using Evolutionary Multivariate Bayseian Prediction Method,Global Journal of Computer Science and Technology: E Network, Web & Security,1-9.

[32] G. Rasitha Banu(2016),Predicting Thyroid Disease using Linear Discriminant Analysis (LDA) Data Mining Technique, Communications on Applied Electronics (CAE),4-6.

[33] Suman Pandey, Anshu Tiwari , Akhilesh Kumar Shrivas ,Vivek Sharma(2015),Thyroid Classification using Ensemble Model with Feature Selection, International Journal of Computer Science and Information Technologies(IJCSIT), 2395-2398.

[34] Wenfeng Song, Shuai Li, Ji Liu, Hong Qin, Bo Zhang, Shuyang Zhang, and Aimin Hao(2018), Multi-task Cascade Convolution Neural Networks for Automatic Thyroid Nodule Detection and Recognition, JOURNAL OF L ATEX CLASS FILES,1-11.

[35] Umar Sidiq, Dr. Syed Mutahar Aaqib, Dr. Rafi Ahmad Khan(2019), Diagnosis of Various Thyroid Ailments using Data Mining Classification Techniques,International Journal of Scientific Research in Computer Science Engineering and Information Technology(IJSRCSEIT), 131-136.